Name: Mark Vincent Ty
Student No: 2018-21871

Machine Problem 2 Written Report:

To quickly view the results, open the link:
https://github.com/markytools/EE214MachineProblems/blob/master/machineprob2.ipynb

1. a.) Using the formula Cov(X, X) = Var(X), we can assume that the variance of the probability distribution of the temperatures each day can be solved using the covariance equation (since i=j). A function called *julytemps* is created with input parameter t for the random variable T. We created a variable called *temp13_10000* that random samples 31 values from a normal distribution with mean=80 and variance=36. This is done 10000 times. We then calculate the probability P[Y >= T] by summing up all Y = “average temperatures” greater than or equal to t.
   b.) In the *julytemps* function, we also return the minimum temperature for each 10000 samples. We then get the samples with the joint distribtion A = {Y <= 82, min Ti >= 72}, and then get the probability. However, there seems to be no samples that hit this event A.

2. (See python code)
   a.) We create and show a Markov chain diagram that detaliates the given Markov state process.
   b.) We then compute the n-step transition probability using the function *markovdisk(n)* with the formula $\mathbf{P}^n$**.**

3. a.) The paper noted that they assumed the observational data was produced from a Gaussian model. A Gaussian model is a good candidate for a prior model when solving the transient probabilities. This is due to the fact as n, the number of data samples, approaches infinity, it can be safe to assume that a Normal distribution could be a possible resulting distribution for the steady state matrix. This is because Gaussian processes are always encountered in real life situations.
   b.) The authors exploited the strength of an HMM model since it is capable of learning the statistical properties extracted from the stock prices. Furthermore, due to the dataset being stochastic in nature, since it is based on the stock value at a certain time interval from the past, the HMM was able to deduce certain patterns in the distribution from which the price datasets where collected.
   c.) A likelihood value for the current day is predicted, based on the current HMM model. From the training dataset, the instances of past data wherein the nearest likelihood are collected. The model parameters (**A, B,** π) are then adjusted to make HMM predict the current price similarly from the past price actions.
   d.) For example, according to the paper, an HMM model was trained between the period of 18 December 2002 and 29 December 2004. To predict the next current price, the HMM needs to acquire the price data produced by the calculated likelihood, which was -9.4544.e.

e.) As shown in some of the figures in the paper, the HMM model was able to show a correlation between the predicted and the actual price. Predicted points were seen clustering in a specific range, near to one another. This makes the Gaussian prior a good assumption. The HMM does produce a good statistical model for the airlines stocks, however in the paper, the authors do not consider the fundamental factors that affect the price of stock, such a the news. This is a very complex scenario and as such, an HMM might find it difficult to converge to the proper probability distribution if it is too complex. However, given the results from this paper, more or less the same results are achievable when transition onto another stock market.

4. a.) Machine learning involves a lot of data, and in order to infer a proper prediction, we need to have the ground truth probability distribution of the data. Sadly, this distribution is unknown to us, and we could only estimate it based on machine learning techniques. This makes Statistics a very important prerequisite of Machine Learning, since we need to learn better architectures, better theories, and better understanding of the underlying statistical structure of the data so that we could derive the best model that can approximate the ground truth probability distribution. Examples of statistical methods applied to machine learning:
    - Problem Framing - exploring data and finding out the problem
    - Data Understanding - summarizing and visualizing data
    - Data Cleaning - removal of outliers to make the data more consistent
    - Data Selection - extracting features from the data that would most likely produce the best resulting score
    - Data Preparation - transformation, encoding, labelling, etc.
    - Model Evaluation - running experiments
    - Model Configuration - using hypothesis testing and estimation
    - Model Selecting - selecting the best model parameters
    - Model Presentation - estimation the best confidence intervals from the network
    - Model Predictions - usage of prediction intervals

    b.) See python code

c.) The Law of Large Numbers applies to any distribution. As n, the total number of samples in the datasets, becomes large, the sample distribution becomes more and more concise and gets close to the population distribution. This can be directly applied to machine learning, wherein a good amount of data is sufficient to create a good model for prediction. The Central Limit theorem states that, as n reaches infinity, the model approximates to a normal distribution. This can be exploited when finding out the sample mean of the data, wherein we could assume that the density can model a Gaussian distribution. This can be used in simple methods such as linear regression.

d.) In hypothesis testing, the correlation can be used to tell whether one variable depends or is associated with another variable. It can describe whether both variables go the same direction, go in opposite, or not at all. Covariance describes the average between the differences of different random variables with their means. It can describe the direction to how the variables relate to one another. It can also show if they are independent or not.

e.) Implementing a Tolerance Interval for a Gaussian distribution introduces a typical uncertainty in our distribution. Once the sample size increases, there is more uncertainty in our data (see python code for sample implementation).